

Claims

What is claimed is:

5

1. A speech reference enrollment method, comprising the steps of:

10

- (a) receiving a first utterance of a vocabulary word;
- (b) extracting a plurality of features from the first utterance;
- (c) receiving a second utterance of the vocabulary word;
- (d) extracting the plurality of features from the second

utterance;

15

(e) determining a first similarity between the plurality of features from the first utterance and the plurality of features from the second utterance;

(f) when the first similarity is less than a predetermined similarity, requesting a user to speak a third utterance of the

20

(g) extracting the plurality of features from the third utterance;

(h) determining a second similarity between the plurality of features from the first utterance and the plurality of features from the third utterance; and

(i) when the second similarity is greater than or equal to the predetermined similarity, forming a reference for the vocabulary word.

5 2. The method of claim 1, further including the steps of:

10 (j) when the second similarity is less than the predetermined similarity, determining a third similarity between the plurality of features from the second utterance and the plurality of features from the third utterance;

 (k) when the third similarity is greater than or equal to the predetermined similarity, forming the reference for the vocabulary word.

15 3. The method of claim 2, further including the steps of:

 (l) when the third similarity is less than the predetermined similarity, returning to step (a).

20 4. The method of claim 1, wherein step (c) further includes the steps of:

 (c1) determining a duration of the second utterance;

25 (c2) when the duration is less than a minimum duration, disregarding the second utterance.

5. The method of claim 1, wherein step (c) further includes the steps of:

- 5 (c1) determining a duration of the second utterance;
(c2) when the duration is greater than a maximum duration, disregarding the second utterance.

6. The method of claim 5, wherein step (c1) further includes the steps of:

- 10 (i) setting an amplitude threshold;
(ii) determining a start time when an input signal exceeds the amplitude threshold;
15 (iii) determining an end time, after the start time, when the input signal is less than the amplitude threshold;
(iv) calculating the duration as a difference between the end time and the start time.

20 7. The method of claim 1. wherein step (d) further includes the steps of:

- (d1) determining an estimate of a number of voiced speech frames;

(d2) when the estimate of the number of voiced speech frames is less than a threshold requesting the user repeat the vocabulary word;

(d3) returning to step (c);

5

8. The method of claim 1, wherein step (a) further includes the steps of:

10

(a1) determining a signal to noise ratio of the first utterance;

(a2) when the signal to noise ratio is less than a predetermined signal to noise ratio, increasing a gain of a voice amplifier.

15

9. The method of claim 8, further including the step of:

(a3) requesting the user repeat the vocabulary word.

20

10. The method of claim 1, wherein step (b) further includes the step of:

(b1) determining an amplitude histogram of the first utterance.

25

11. A speech reference enrollment method, comprising the steps of:

- 5 (a) requesting a user speak a vocabulary word;
(b) detecting a first utterance;
(c) requesting the user speak the vocabulary word;
(d) detecting a second utterance;
(e) determining a first similarity between the first utterance
10 and the second utterance;
(f) when the first similarity is less than a predetermined similarity, requesting the user speak the vocabulary word;
(g) detecting a third utterance;
(h) determining a second similarity between the first utterance
15 and the third utterance; and
(i) when the second similarity is greater than or equal to the predetermined similarity, creating a reference.

12. The method of claim 11, further including the steps of:

- 20 (j) determining a third similarity between the second utterance and the third utterance;
(k) when the third similarity is greater than or equal to the predetermined similarity, creating the reference.

13. The method of claim 12, further including the steps of:

(1) when the third similarity is less than the predetermined similarity, returning to step (a).

5

14. The method of claim 11, wherein step (b) further includes the steps of:

(b1) determining if the first utterance exceeds an amplitude threshold within a timeout period;

10

(b2) when the first utterance does not exceed the amplitude threshold within the timeout period, returning to step (a).

15

15. The method of claim 11, wherein step (b) further includes the steps of:

(b1) determining an estimate of a number of voiced speech frames;

20

(b2) when the number of voiced speech frames is less than a predetermined number of voiced speech frames, returning to step (a).

16. The method of claim 11, wherein step (b) further includes the steps of:

25

(b1) determining a duration of the first utterance;
(b2) when the duration is less than a minimum duration,
returning to step (a);
(b3) when the duration is greater than a maximum
5 duration, returning to step (a).

17. A computer readable storage medium containing
computer readable instructions that when executed by a computer
10 performs the following steps:

(a) requesting a user speak a vocabulary word;
(b) receiving a first digitized utterance;
(c) extracting a plurality of features from the first digitized
15 utterance;
(d) requesting the user speak the vocabulary word;
(e) receiving a second digitized utterance of the vocabulary
word;
(f) extracting the plurality of features from the second digitized
20 utterance;
(g) determining a first similarity between the plurality of
features from the first digitized utterance and the plurality of
features from the second digitized utterance;
(h) when the first similarity is less than a predetermined
25 similarity, requesting the user to speak a third utterance of the
vocabulary word;

(i) extracting the plurality of features from a third digitized utterance;

(j) determining a second similarity between the plurality of features from the first digitized utterance and the plurality of features from the third digitized utterance; and

(k) when the second similarity is greater than or equal to the predetermined similarity, forming a reference for the vocabulary word.

10 18. The computer readable storage medium of claim 17, further executing the steps of:

15 (l) when the second similarity is less than the predetermined similarity, determining a third similarity between the plurality of features from the second digitized utterance and the plurality of features from the third digitized utterance;

(m) when the third similarity is greater than or equal to the predetermined similarity, forming the reference for the vocabulary word.

20 19. The computer readable storage medium of claim 18, further executing the steps of:

25 (n) when the third similarity is less than the predetermined similarity, returning to step (a).

20. The computer readable storage medium of claim 17,
wherein step (c) further includes the steps of:

- 5 (c1) determining a signal to noise ratio;
(c2) when the signal to noise ratio is less than a
predetermined signal to noise ratio, going to step (a);
(c3) determining if an amplifier gain is saturated;
(c4) when the amplifier gain is saturated, going to step
10 (a).

21. The computer readable storage medium of claim 20,
wherein step (c2) further includes the step of increasing a gain of an
amplifier.

15 22. The computer readable storage medium of claim 20,
wherein step (c4) further includes the step of decreasing a gain of an
amplifier.